DOCUMENT RESUME

ED 252 544 TM 840 758

AUTHOR Taylor, Marcia B; Porterfield, William D.
TITLE Research Data on the Measure of Epistemological Reflection: Measuring the Perry Scheme.
PUB DATE 84
NOTE 16p.; Paper presented at the Annual Meeting of the American College Personnel Association (Baltimore, MD, April 1984). For a related document, see ED 235 438.
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Cognitive Development; *Cognitive Tests; *Epistemology; Higher Education; Interrater Reliability; Research Needs; Scoring; Test Construction; Test Manuals; Test Reliability; *Test Validity
IDENTIFIERS *Measure of Epistemological Reflection; Perry Scheme of Intellectual Ethical Development

ABSTRACT

This paper describes the Measure of Epistemological Reflection (MER), an instrument to assess cognitive developmental level according to the Perry scheme of intellectual and ethical development. It contains sets of questions for each of the six cognitive domains: decision making, learner role, instructor role in the learning process, peer role in the learning process, evaluation systems, and the nature of truth or reality. The first question focuses the respondent's thinking on the domain while followup questions probe for the justifications for the respondent's thinking. The rating manual has the same format, with specific response examples to reduce inference in the rating process. The MER was administered to four samples of undergraduate and graduate students (n=621). Reliability was explored through interrater agreement (exact agreement percentage on domain ratings and modal total protocal ratings, plus correlation of continuous protocol ratings) and internal consistency (one domain rating's consistency with total protocol ratings across subjects). Validity was investigated three ways: (1) content validity by the match of test domains to the Perry scheme; (2) concurrent validity by comparisons to Measurement of Intellectual Development results; and (3) construct validity by analysis of variance by level of education. (BS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED252544

RESEARCH DATA ON THE MEASURE OF EPISTEMOLOGICAL REFLECTION: MEASURING THE PERRY SCHEME

Presented at the American College Personnel Association's National Convention

Baltimore, Maryland, 1984

Taylor, Marcia B.
Porterfield, William D.

© 1984

TM 810 758

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

M. B Taylor

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

RESEARCH DATA ON THE MEASURE OF EPISTEMOLOGICAL REFLECTION: MEASUREMENT ON THE PERRY SCHEME

Presented at the American College Personnel Association's National Convention, Baltimore, Maryland, April, 1984

INTRODUCTION

The Perry scheme of intellectual and ethical development (Perry, 1970; 1981) has provided student personnel practitioners, teachers, educational administrators and researchers with a framework for understanding the cognitive structural changes in the way individuals make meaning of their experiences. The theory has been translated into practical models in such areas as curriculum instructional design (Knefelkamp, 1974; Widick, 1975; Copes, 1974), faculty consultation (Parker, 1978), career counseling (Slepitza and Knefelkamp, 1976) and employee supervision and training (Note 1). Practitioners and researchers continue to find ways to apply the theory in a variety of settings in higher education. The continued use of the Perry scheme in higher education settings (both research and practice) highlights the need for valid and reliable assessment instruments. The development of the Measure of Epistemological Reflection (Note 2) represents an effort to provide researchers and practitioners with a paper and pencil instrument to assess cognitive developmental level on the Perry scheme. A description of the instrument and its properties, the initial validation study, subsequent studies and future possibilities are outlined below.

THE INSTRUMENT AND ITS PROPERTIES

The Measure of Epistemological Reflection (MER) consists of six sets of questions. Each set addresses a content area relevant to the Perry

scheme. The six domains are: decision making, role of the learner, role of the instructor in the learning process, role of peers in the learning process, evaluation systems, and the nature of truth or reality. Each set of questions begins with a general question to focus the respondent's thinking on the domain. Follow-up probes serve to elicit the respondent's justification for his/her thinking. The respondent's justifications for their responses are then compared to those in a rating manual, and assigned a modal stage (positions one through five) based on the rating manual. Each of the six domain areas in the instrument receives a modal stage rating, and the entire instrument (protocol) is assigned a rating based on the frequency of stage responses. The instrument is suitable for group administration, and takes approximately one hour to complete.

The construction of the MER was based on a number of foundations established by other researchers in instrument development. The work of Loevinger and Wessler (1970), Kohlberg (1981) and Gibbs (1982) provided direction for the current instrument in terms of format and rating procedures.

Within the individual, various strands of development are taking place simultaneously. This creates a difficulty for assessment inasmuch as assessment procedures must attempt to identify particular strands of development as focal points. The six domains established for the MER were an attempt to specifically define development on the Perry scheme in order to distinguish this strand from other strands of development. A second major difficulty emerges when indicators of development that appear in earlier stages reappear in more complex form in later stages. Descriptions of qualitative differences across Positions One through Five in each domain served to clarify distinctions between these indicators.

Another potential difficulty in the assessment of cognitive structures is the question format. Previous instruments for the Perry scheme have included sentence completion tasks and essay questions. The difficulty with these two formats is that questions may not be specific enough to guide the respondent into a cognitive domain that is ratable and, with respect to essay questions, the respondent may selectively choose to answer some parts of the essay and not others. To avoid these difficulties, and to increase the probability that respondents answers are within a domain of interest to assessment, the MER employs a format which includes an initial question which asks for an opinion about an issue(s) in the cognitive domain. This question is then followed by a number of probe questions which ask the respondent to justify or explicate his/her reasons for the opinion expressed in response to the first question in the series.

A third property of the MER is related to the question format described above. In rating protocols, the rater is concerned with the respondent's justification for an opinion as opposed to the opinion itself. Since Perry's scheme of development describes changes in the individual's reasoning structure, it is important to separate content and justification in rating responses. The probe questions within each domain on the MER serve to elicit the respondent's justification for a particular position. It is the respondent's justification, versus the opinion, that serves as the rater's cue to assigning a stage to the data.

A fourth concern in the construction of an assessment instrument is that the instrument's stimuli elicit adequate data for rating purposes. Since the format of the MER focuses the respondent's thinking, asks for justification, and includes several probe questions there is increased

probability that the respondent will generate enough data for the rater to assign a modal stage to the response.

Rating manuals have been a source of continued concern in assessment for the Perry scheme. Manuals to date have consisted of broad descriptions of the scheme and cues to assist the rater in assigning stage positions to the protocols. The MER uses a rating manual for each domain. Each section of the manual contains a position description relevant to the particular domain. This overall description is followed by a list of reasoning structures, or justifications, which would most likely be employed by a respondent at the particular position on the Perry scheme. The reasoning structures are followed by examples from protocols. The intention of the extended manual is to reduce the amount of inference necessary to rate a particular response. The rater is asked to identify the modal stage and the reasoning structure for each domain within a protocol.

Loevinger (1976) and Gibbs (1982) note that manual construction for such instruments is an ongoing process. The first step in creating such a manual involves describing the theory in such a way that it is useful to the rater in working with data elicited by the instrument. Next, instruments are piloted and rated based on the preliminary manual. As the data base continues to build, the manual may be revised based on the data. In the case of the MER, additional reasoning structures were added to the manual through an empirical verification process (Loevinger, 1970). This process involved dividing the original sample in half, rating the first half with the preliminary manual, generating reasoning structures from the first half of the data and rating the second half of the data with the revised manual. These reasoning structures provided

additional insight into the nuances of a particular position of development.

Aside from the content and format of a rating manual, particular attention must be given to a detailed set of rating instructions. These rating instructions guide the raters, and insure that data are being rated using consistent rules. With the MER, the rater is asked to read all of a particular domain response and to decide, based on the overall position descriptions, where the data seems to "fit" best. The rater then surveys the reasoning structures under the position description to determine the best match. While a particular domain may contain evidence of more than one position, the rater is asked to identify the modal position within the domain. Each domain is assigned a position and a reasoning structure from the manual, and total protocol ratings (TPR) are determined through calculation of position frequencies.

In summary, the MER is constructed on six cognitive domains relevant to the Perry scheme. Each domain is represented in the instrument by a set of questions. The first question focuses the respondent's thinking within the domain. A series of probe questions follow, which serve to elicit the respondent's justifications or reasoning structures in each domain. The rating manual is based on the same format as the instrument, and contains examples which are matched to the data under consideration. This increased specificity in the stimuli and rating process is intended to reduce inference in the rating process.

RESULTS

The MER has been administered to four samples totalling 621 subjects. The original sample (N=155) was divided in half to form a

derivation sample and a cross-validation sample which were rated with a preliminary and revised manual respectively (Taylor, 1983). The results of this initial validation study (Taylor, 1983) indicated sufficient reliability and validity data to warrant continued studies of the MER with other samples. In addition to these two subsamples the MER has been administered to a sample of undergraduate residence hall staff (N=165), undergraduate and graduate students enrolled in a college of social work (N=121), and undergraduate and graduate students in a teacher education program (N=180).

Reliability

Reliability of the MER was explored through interrater agreement and internal consistency. Three levels of interrater agreement were considered: percentage of exact agreement on domain ratings, percentage of exact agreement on modal total protocol ratings (TPR) and correlation of continuous total protocol ratings. Agreement on domain ratings was utilized to explore areas of the manual that required additional specificity. This level of agreement also served as the most stringent assessment of interrater agreement since a domain constitutes one sixth of the instrument. The latter two levels provided an assessment of agreement on a subject's position as revealed by the entire protocol.

The rating procedure involved at least two raters independently rating the protocols for a sample. Upon completion, levels of agreement were established and all disagreements on domains were discussed to reach a compromise rating. This compromise rating was utilized in the original sample for comparison with self-trained raters (Taylor, 1983) and in remaining samples to provide subjects developmental levels for use in the

studies. The compromise rating also serves as the basis for tests of validity. The authors of the MER rated all the samples mentioned. Given the discussions in the compromise phase, it was expected that agreement would increase in subsequent rating. In addition, two revisions of the manual, first on the basis of empirical data in the derivation sample and a second time prior to the teacher education sample, served to clarify distinctions among reasoning structures and was expected to increase overall agreement. Thus, the discussion that follows addresses agreement within each sample and across samples in the order in which they were rated.

Exact agreement on domain ratings for the derivation sample (N=84) ranged from 41.7% to 64.3%. The lowest agreement was in domain one (decision making), which generates the most diversity in responses. Domain four (role of peers) revealed 50% agreement while the remaining four domains revealed 58.3% to 64.3% agreement. While the range of agreement in the cross validation sample (N=71) was similar, exact agreement in domains one and four increased to 54.9% and 57.7% respectively. The increased specificity of the manual in these two areas seemed to reduce inference in rating. Domains two (role of the learner) and three (role of the instructor), both showing agreement of 64.3% in the derivation sample, decreased to 46.5% and 54.9% respectively. The manual in these areas was more specific for the cross-validation sample but did not sufficiently clarify subtle distinctions between Positions Two and Three discovered in the derivation sample. Post-rating discussion of these distinctions proved helpful in rating the residence hall sample. Exact agreement was 53.6% in domain two and 67.5% in domain three. Agreement in the remaining domains increased slightly, with domain one

remaining at 54%.

The social work sample consisted of juniors, seniors and graduate students in contrast to mostly sophomore and juniors in the residence hall sample. Subsequently, the ratings of the residence hall sample were in Positions Two and Three, while those in the social work sample were generally Positions Three and Four. Thus, while previous discussion of distinctions between Positions Two and Three helped in rating the residence hall sample, the social work sample called for discriminating between Positions Three and Four. This situation contributed to the decreased agreement in the social work sample evidenced in domains one (49.6%), three (59.5%), four (50.4%) and six (52.9%). A second revision of the manual following the social work sample served to clarify distinctions across all five Perry positions on the basis of all empirical data available. The percentage of exact agreement on domains in the teacher education sample reflects the increase in manual specificity. Agreement in domains one and five remained similar to previous samples but the remaining domains revealed increases from 9% to 16% over the social work sample. The range of agreement in the teacher education sample, with the exception of domain one at 47.8%, was 60% to 75%. It should be noted that, in addition to exact agreement, agreement within one stage was calculated. With the exception of domain one, agreements within one stage brought agreement on domain ratings to the 90% range.

Two procedures were developed for acquiring total protocol ratings (TPR). The first involved determination of the modal position of each subject as evidenced by the six domain ratings. Although Perry's scheme indicates that a person could be in more than one position simultaneously, the theory also indicates that it is logical to assume that each person

has a level of modal functioning. Thus, transition was also accounted for in the assignment of TPR's by including positions revealed by two or more domains (Taylor, 1983). The second procedure involved averaging the six domain ratings to acquire a continuous TPR that reflected all domain ratings.

Interrater reliability on modal TPR's was established by percentage of exact agreement. Exact agreement on modal TPR's was 59.5% for the derivation sample, increasing as expected on the cross-validation sample (73.2%). Agreement on the residence hall sample remained similar at 72.1%. Agreement decreased to 63.6% on the social work sample due to the decreases in domain agreement mentioned previously. Agreement again increased on the teacher education sample to 74.4%. Agreement on modal TPR's within one stage brought total agreement to 96.4% to 100% across samples.

Pearson product moment correlation coefficients calculated for continuous TPR's show interrater reliability similar to that on modal TPR's. The highest correlations were found in the derivation and cross-validation samples, with .82 and .78 respectively ($p < .0001$). The correlation dropped to .62 ($p < .0001$) for the residence hall sample but was affected by the range of scores being restricted to 2.6 to 3.2. The social work sample revealed a correlation of .69 ($p < .0001$). This correlation was affected by lower agreement noted earlier in this sample, as well as a narrow range of scores given the makeup of the population. The teacher education sample, which like the derivation and cross-validation samples included a wider range of educational levels, revealed a correlation of .73 ($p < .0001$). Thus the correlation for all protocols across samples was .79 ($p < .0001$). The collective results on interrater

reliability indicate that while continued clarification in some areas of the manual is needed, raters can achieve an acceptable level of consistency using the current manual.

Internal consistency

Internal consistency was evaluated with Cronbach's alpha coefficient which assessed the degree to which ratings in one domain were consistent with total protocol ratings across subjects. The alpha for the original sample (derivation and cross-validation combined) was .76, a moderately high level of consistency given that subjects could theoretically be expected to be in some transition position as well as a modal position. Alpha coefficients for subsequent samples were .63 (residence hall staff sample), .60 (social work sample) and .65 (teacher education sample). The coefficient for all samples was .74.

Validity

Validity of the MER was investigated in three ways. First, the MER was judged to have content validity since the domains included in the MER were generated from the Perry scheme and subsequent research, and since the six domains cover the range of the content on the scheme. Concurrent validity was investigated in the derivation samples through comparison to the Measure of Intellectual Development (Note 3). Results were inconclusive due to a restricted range of ratings on the Measure of Intellectual Development (Taylor, 1983). Sample and overall means of freshmen, seniors and graduate students supported concurrent validity of the MER due to their similarity to means of those groups found in similar research (Note 3; Note 4; Meyer, 1977).

Construct validity of the MER was examined through an analysis of variance by level of education. An analysis of variance of the original sample (derivation and cross-validation combined) indicated significant differences by level of education ($p < .0001$). Scheffe post hoc multiple comparisons revealed significant differences ($p < .05$) between freshmen, seniors and graduate students. No significant differences emerged on the basis of sex. An analysis of variance of the residence hall staff sample revealed no differences, again due to the restricted range of educational levels in that population. Significant differences ($p < .0001$) reappeared in the social work sample. Due to the conservative nature of the Scheffe test (Kennedy, 1977), the Student Newman Keuls post hoc multiple comparison was used to determine where the differences existed. The comparisons indicated significant differences ($p < .05$) between Ph.D. candidates, second year masters candidates and the remaining subjects who were juniors, seniors and first year masters candidates. Analysis of variance of the teacher education sample also resulted in significant differences ($p < .0001$). Student Newman Keuls comparisons ($p < .05$) indicated significant differences between Ph.D. candidates, first year masters candidates and the remaining subjects who were seniors, juniors, sophomores and one freshman. Since only one freshman was included in the teacher education sample, this may account for the lack of difference between seniors and freshmen. The group means increased with level of education in each sample. Finally, an analysis of variance of the total sample (N=621) revealed differences by level of education ($p < .0001$). Student Newman Keuls comparisons ($p < .05$) indicated that Ph.D. candidates and second year masters candidates differed from first year masters candidates and seniors. This latter group in turn differed from juniors,

sophomores and freshmen. These results taken collectively provide strong evidence that the MER accurately measures development on the Perry scheme.

FUTURE RESEARCH DIRECTIONS

Current research findings warrant further investigation of the MER in the following areas. First, research must be conducted to determine correlations between the MER and interview data. Second, longitudinal studies are essential to confirm construct validity. Third, the MER should be administered to diverse populations in terms of age and socioeconomic background. Fourth, intervention studies would be helpful in reinforcing the validity of the MER. Fifth, a strategy is currently being devised to determine the feasibility of a self-training packet which would further the use of the MER in research and practical settings.

NOTES

1. Ricci, J.P.; Porterfield, W.D. and Piper, T.D. "Developmental Supervision: Renewing a Commitment to Employee Development." Paper presented at the American College Personnel Association National Convention, Baltimore, Maryland, 1984.

2. Taylor, M.B. and Porterfield, W.D. "The Measure of Epistemological Reflection." 1982.

3. Moore, W. "Measure of Intellectual Development." Prepublication Draft, Center for the Application of Developmental Instruction, University of Maryland, 1982.

4. Rodgers, R.F. Unpublished studies. The Ohio State University, 1975-1979.

BIBLIOGRAPHY

Copes, L. "Teaching models for College Mathematics." Doctoral dissertation, Syracuse University, 1974.

Gibbs, J.C. & Widaman, K.F. Social Intelligence: Measuring the Development of Sociomoral Reflection. Englewood Cliffs, California: Prentice Hall, Inc., 1982.

Kennedy, J.J. An Introduction to the Design and Analysis of Experiments in Education and Psychology. Washington, D.C.: University Press of America, 1977.

Knefelkamp, L.L. "Developmental Instruction: Fostering Intellectual and Personal Growth of College Students." Doctoral dissertation, University of Minnesota, 1974.

Kohlberg, L. The Philosophy of Moral Development. San Francisco: Harper and Row, 1981.

Loevinger, J. & Wessler, R. Measuring Ego Development, Volume One. San Francisco: Jossey-Bass, Inc., 1970.

Meyer, P. "Intellectual Development: Analysis of Religious Content." The Counseling Psychologist, 1972, 6,4: 47-50.

Parker, C.A. "Individualized Approach to Improving Instruction." National Association of College Teachers of Agriculture Journal, 1978, 22, 14-28.

Perry, W.G. Forms of Intellectual and Ethical Development in the College Years: A Scheme. New York: Holt, Rhinehart and Winston, 1970.

Perry, W.G. "Cognitive and Ethical Growth: The Making of Meaning." In A. Chickering, The Modern American College. San Francisco: Jossey-Bass, 1981.

Slepitza, R. & Knefelkamp, L.L. "A Cognitive Developmental Model of Career Development: An Adaptation of the Perry Scheme." The Counseling Psychologist, 1976, 6,3: 53-58.

Taylor, M.B. "The Development of the Measure of Epistemological Reflection." Doctoral dissertation, The Ohio State University, 1983.

Widick, C. "An Evaluation of Developmental Instruction in a University Setting." Doctoral dissertation, University of Minnesota, 1975.